MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

Memo No. 325

December 1974

THE LOW-LEVEL SYMBOLIC REPRESENTATION OF INTENSITY CHANGES IN AN IMAGE

by

David Marr

ABSTRACT

A family of symbols is defined by which much of the useful information in an image may be represented, and its choice is justified. The family includes symbols for the various commonly occuring intensity profiles that are associated with the edges of objects, and symbols for the gradual luminance changes that provide clues about a surface's shape. It is shown that these descriptors may readily be computed from measurements similar to those made by simple cells in the visual cortex of the cat. The methods that are described have been implemented, and examples are shown of their application to natural images.

Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract number N00014-70-A-0362-0005.

Summary

1. It is shown that at least three main kinds of intensity change are commonly to be found in an image. These are (a) step changes in intensity, such as exist at well-illuminated object boundaries against a dark background; (b) step changes in intensity gradient, such as can exist on the shaded side of a curved object against a dark background; and (c) gradual changes in intensity over a surface, due to the combined effects of the surface's shape, and the prevailing illumination.

2. The characteristics of each kind of intensity edge are noted, and symbols for each are defined. It is shown how the different kinds of intensity edge may be recognised by an orientation-dependent analysis of an image. They may most straightforwardly be computed from edge- and bar-mask convolutions with the image, since these measure appropriate approximations to the first and second directional derivatives of intensity.

3. In addition to the classification of intensity changes, additional descriptors are defined to represent properties of "strength" and "fuzziness". The computation of these measures requires the comparason of edge and bar mask convolutions made with masks of two or more different sizes, and a detailed account of methods for doing this is given.

4. The methods that are described have been implemented using serial algorithms on a conventional computer, and examples are shown of their use on selected images. These methods will be successful provided that (a) the image resolution is generous compared with the distance between intensity changes, and (b) that the image is examined at the appropriate scale. Methods for discovering the appropriate scale are given. It is noted that advanced mammalian visual systems are designed in a way that would enable these conditions to be satisfied.

5. Parallel algorithms are available for many parts of the parsing process.

6. The intensity distribution that one would infer from the symbolic representation of an image frequently differs from the true intensity distribution. These anomalies may illuminate the cases where our own perception of these distributions is similarly, and usefully, deceiving.

Introduction

It was argued elsewhere (Marr 1974a) that the object of the first stage in a general-purpose vision system should be to compute a low-level description of the intensity changes that occur in an image, from suitably chosen measurements made upon it. The low-level description should consist of a set of assertions and modifiers, that employ appropriately chosen atomic symbols drawn from a vocabulary of adequate power. It was also pointed out that the measurements from which the description is computed should be orientation dependent in a simple way. This is because the difficulty of computing a symbolic low-level description is related to the structure of the inverse transform of the original measurement: if the inverse transform has a complex dependence on boundary conditions, so must the computation of the description.

This article examines the kinds of intensity change that commonly exist in natural images, defines a vocabulary of low-level symbols in terms of which they may be described, and gives methods by which this representation may be computed. The present vocabulary was based on a combination of intrinsically important computational criteria that arise in the deciphering of bar- and edge-mask convolution profiles, together with the requirement that the method describes all of the changes that one can see oneself in the images (apart from features knowingly omitted, like specularities). In some respects, the system set out here is more complete than our own perception of an intensity array, but there may be others in which it is less. One must however start somewhere, and if the

Figure 1. 1a shows two of the simple paper surfaces that provided some of the images shown in later figures. The camera was placed directly over the surfaces, which were lit from the side. The surfaces may be identified by the profile number that appears beside them. Figure 1b shows examples of edge and bar masks, with the weights that were used. Notice that the weights have been chosen to take account of the rectangular tesselation of the image.

(a)

(b)

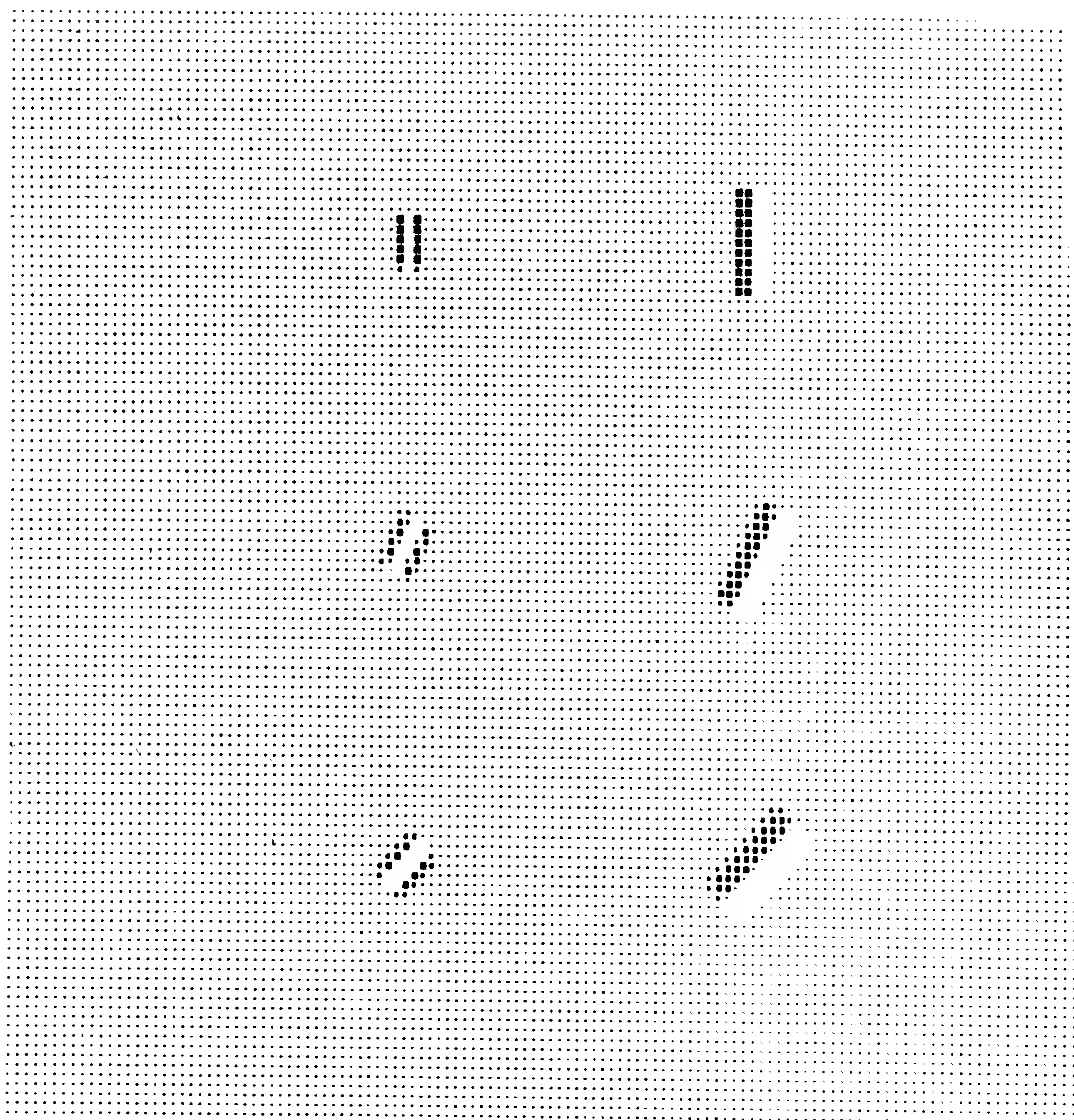


FIGURE 1

terms defined here turn out not to be quite precise enough for very fine discriminations, one can refine them slightly while preserving the overall method. This question will be raised again in subsequent articles, as we follow the manipulation of the low-level description described here up to its representation in terms of higher level predicates through which we are accustomed to perceiving the world.

Methods

The pictures were taken using a Telemation TMC-2100 television camera. When set in log mode, and adjusted appropriately, this camera delivers a signal corresponding to 2/3 log intensity over a reasonable range. Care was taken to ensure that the scenes photographed fell within saturation levels of the camera. Image output was provided by a DEC-340 display unit, and by a xerographic line-printer on-line to the central computer installation based on a DEC PDP-10. The photographs of these images that appear here are necessarily inaccurate representations of the underlying distributions, but they preserve and sometimes enhance the qualitative features with which the programs are concerned. The images were created by bending and folding pieces of white paper, as illustrated in figure 1a, and by photographing various common objects. The illumination consisted of a roughly uniform component due to diffuse overhead lighting, together with a local source provided by a standard desk lamp, which was responsible for the shadows that appear in the images. The grey-level dynamic range of the camera was 8 bits, and that of a typical picture exceeded 7.

Edge- and bar-mask convolutions

In the early stages of this study, I had the pre-conception that step-changes were the only important kind of intensity change in an image. (We shall not be concerned here with changes due to motion, or to disparity). Both edge-shaped and bar-shaped masks (figure 1b) are equally good at detecting this kind of change, and it was therefore a disturbing puzzle that the cat's visual cortex contains both kinds of simple cell receptive field (Hubel & Wiesel 1962). One possibility was that one type of mask should somehow be closely related to the detection of bars in the image, and the other, to the detection of edges. This cannot be the case, however, because bar-mask convolutions are very different from assertions about the presence of bars in an image (see Marr 1974a).

An edge-shaped mask can be viewed as signalling an approximation at a certain scale to the first directional derivative of intensity at a point; and a bar-shaped mask as signalling either an approximation to the second directional derivative, or to the difference between the left and right first derivatives. Once one realises this, and also that pure step-changes in intensity are only one of a number of kinds of intensity change, it becomes very reasonable to view the convolutions as measuring the first and second derivatives of intensity. These two measurements convey information about the way intensity is changing at a point (higher derivatives being much less interesting); and they have the property of orientation selectivity that is virtually required of the measurements

from which the low-level description is to be computed. Finally, it is worth pointing out that the units in which the derivatives should be expressed are units of contrast ((1/I)dI/dx), for intensity I); or equivalently the gradient of log intensity (since d/dx (log I) = (1/I)dI/dx). This is because the gradients on two surfaces, that have the same illumination but different reflectances, have the same value in these units, but different values in pure intensity gradient units.

The size of an edge- or bar-mask is characterised by the width of one of its constituent panels. This is called the panel-width, and varies between 1 and 64 in the figures that accompany this article. The length of a mask is typically 4 or 5 times its width: the reason for not having it shorter is that inter-orientation cross-talk is small for masks of this length. For a given mask, a convolution profile may be obtained by computing the mask response across the image along a line whose orientation is perpendicular to the principal orientation of the mask. Much of the discussion below concerns the peaks and slopes that occur in such a profile: an example of one appears in figure 3.

The process of computing edge- and bar-mask convolutions is of some interest. The visual cortex of the cat performs the convolution directly, using what are effectively hard-wired masks scattered at all positions and orientations over the visual field. Direct simulation of this on a serial machine is very inefficient: it is much faster to regard the operation as a convolution, and to use the Fast Fourier Transform (FFT) algorithm (Cooley & Tukey 1965) to reduce the convolution to a multiplication. In the implementation that is described here, the

convolution is taken over an array of log intensity values. This produces a measurement that approximates the local contrast gradient, to whose logarithm simple cells in the cat appear to be sensitive (Maffei & Fiorentini 1973 figure 8).

Parsing the results of the convolutions

What information should one extract from an edge- or bar-mask convolution such as that shown in figure 2? There are two broad options: one can either continue transforming the image - for example by remaining in the spatial frequency domain and applying various convolutions to the whole of the data; or one can extract a few simple measurements like the position and size of the peaks in a profile, perhaps adding a simple descriptor of the sharpness of the peak, and parse these measurements into symbolic assertions about the data. The methods described below take the second approach, and part of their justification is that they work acceptably. But it is important to be aware of the issues that lie behind the choice, so I include a brief discussion of them here.

Whenever one chooses to make a transformation of a piece of data, (for example the Fourier transform of an image), one becomes committed to a notion of similarity that is associated with that transform. In the case of a Fourier transform, the similarity is defined by some metric in the frequency domain. It is necessary to ask whether the particular choice of similarity that the transform introduces is appropriate for the given application. Experience with line-finding programs shows, for

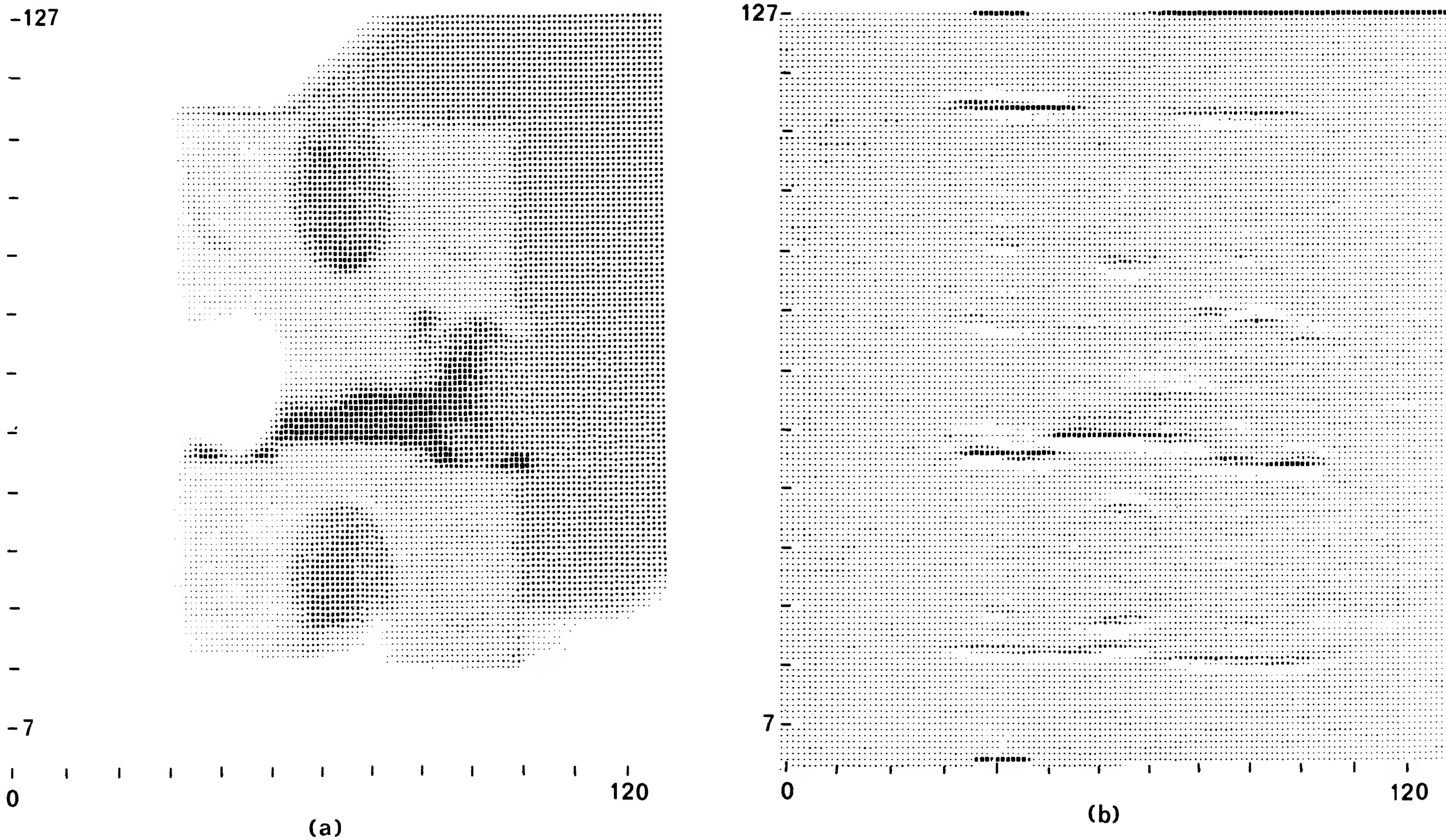


FIGURE 2

Figure 2. To illustrate the results of convolving a bar mask with an image, a picture of one end of a building brick from a Fischer-Technik children's construction kit (2a), has been convolved with a bar mask of panel-width 2 and length 11. The orientation of the mask was horizontal, and the result has been represented as an image in 2b. Zero corresponds in figure 2b to regions of medium intensity.

example, that algorithms based on Fourier style detection methods fail to find many of the interesting lines in an image (B.K.P.Horn personal communication): so that even at the very lowest level, metrics based rigidly on spatial frequency spectra fail to supply the appropriate measures of similarity. I have tried various Fourier techniques for extracting peaks from bar- and edge-mask convolution profiles, but they are too sensitive to the exact shape of the peaks to be useful. Whether a peak is there or not, its position, size, and possibly its thickness, seem to be the important factors.

The use of such qualitative features as these rests upon other assumptions which, if violated, will cause methods that rely on them to produce nonsense. The assumptions are roughly equivalent to the assumption that the boundary conditions, for the local restriction of the inverse transform, may be ignored (Marr 1974a). We may refer to this as the <u>isolation condition</u>. The isolation condition is violated if two edges in the image are so close together that the corresponding peaks in the convolution profile interfere. To allow for this, one has to apply some de-smearing technique before an accurate assessment of the peaks may be made. De-smearing is itself a transformation, however, and therefore brings with it problems related to its own similarity and stability characteristics. One does better to avoid such operations if possible.

In the human visual system, the receptors are spaced at a distance of 20" to 35" apart (see e.g. Cornsweet 1970 p356). Complex patterns that cover a small number of receptors are not well resolved by us, and it seems that we perform well only in those cases where the

Figure 3. 3c shows the intensity distributions of an edge, a wide bar, and a thin bar. These intensity distributions have been convolved with edge and bar masks, whose panel widths equalled the width of the wide bar. The results for the edge mask appear in 3a, and those for the bar mask appear in 3b. Figure 3d shows the bar-mask convolution of an image in which two wide bars are separated by the width of those bars. The panel-width in 3d equals the width of the bars.

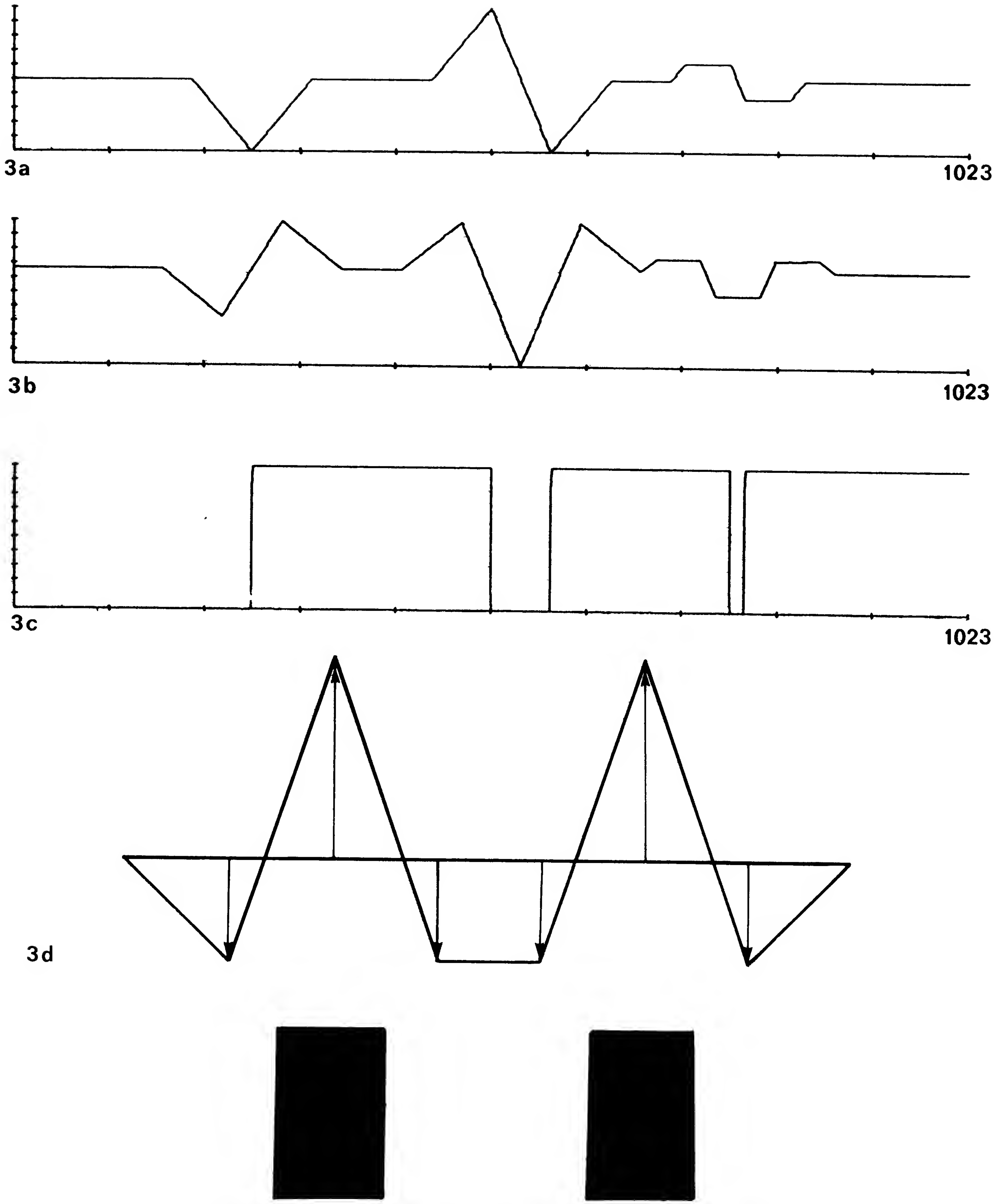


FIGURE 3

details in the image are sufficiently separated for the highest resolution masks to satisfy the isolation criterion. Nevertheless, the study of images in which the detail is very densely distributed is of interest, and I report some experience with them elsewhere (Marr 1974b). In a real-time environment, however, the safest strategy would be to take a closer look if more detail is required, because the necessary computations turn out to be somewhat delicate.

Isolated step changes in intensity

In order to define the parsing process precisely, let us examine the characteristics of various kinds of intensity change, starting with the simplest. Figure 3c shows the intensity distribution of an edge at x=256, a wide bar (x = 512 to 576), and a thin bar (x = 768 to 784). The bar- and edge-mask convolutions with this intensity distribution, for a panel-width of 64, are given as figures 3b and 3a. The salient characteristics of the convolution profiles are as follows:

Sharp edge: this gives rise to a single, sharp peak in the edge-mask convolution. The half-width of the peak is d, the panel-width in the underlying mask. The bar-mask convolution shows a positive and a negative peak, separated by a distance d, and which decline linearly to zero a further distance d out to the sides.

Sharp bar: provided that the width of the bar exceeds 2d, the bar appears as two quite separate edge-mask responses. For narrower bars, the two peaks start to interfere, and their apparent amplitude diminishes

linearly with the width of the bar. Similar observations hold for the bar-mask responses; here, interference starts at a bar width of 3d, and the peak response (the "typical" bar response that is shown in the figure)occurs at a bar width of d, equal to the panel-width.

Lines: When the width of a bar is smaller than the smallest panel-width in use, the bar- and edge-mask convolutions have a cut-off appearance (figures 3a and 3b, around x=770). The position of the underlying line is determined, but the characteristics of the intensity changes at its edges are not.

Finding peaks in a profile

In order to diagnose the presence of a sharp edge (or of anything else) in an image from the characteristics of the peaks in various bar- and edge-mask profiles, those peaks must first be found. This process is of some interest in its own right. First, we define a possible-peak to be a local maximum whose value is positive, or a local minimum whose value is negative. This criterion is extremely liberal: it allows small local bumps to be called possible-peaks, even if they are close together and about the same size. The possible-peaks in profiles from two sizes of mask are then matched, and possible-peaks that occur in both profiles are called peaks. This point is of some importance, because when one looks at a profile from a very small mask, it is usually not possible to state which of the small peaks in it are important and which are due to noise. Combining the information from two (or more) mask-sizes provides a method of peak-detection that is much more sensitive than methods based on only

one profile.

Sharp edges

Sharp edges may be reliably detected by looking for sharp peaks in the edge-mask convolution, or by looking in the bar-mask convolution for two peaks, of the same amplitude but opposite sign, separated by the panel-width d. The advantage of the second method is that bar-masks suffer less from inter-orientation cross-talk than edge-masks: but the disadvantage is that the resolving power for two close edges is inferior unless de-smearing is used. In practise, I have used the first technique with a stringent criterion for sharpness to extract all of the really obvious sharp or slightly fuzzy edges. This gives the program more room to manoeuvre when studying the more difficult cases. The sharpness criterion that is applied in my present implementation is that at a distance d/2 away on at least one side, the value should not exceed 0.55 that of the peak; and that on the other side, it should not exceed 0.8 of the peak. If the latter condition is violated, but satisfied by the point d/2 further along, the edge is described as slightly fuzzy at that resolution. In general, sharpness can better be determined by comparing the amplitudes of the peaks in profiles from masks of different sizes (see below).

Diagnostic procedures from bar-mask convolutions are somewhat more complex, and are dealt with later.

Fuzziness

Pure step-changes in intensity are comparatively rare in natural images. If the change is spread out over a small distance, the edge appears to be fuzzy. As one would expect from the spatial frequency spectrum of such an edge, the small masks give relatively less response to fuzzy edges. For example, consider two edge-masks M and N, where the panel-width of N is half that of M. If their responses are normalised so that the response of each to a step function is 1, the response of M to a linear slope is twice that of N, because the mean separation of the panels is doubled. Hence, by comparing the relative sizes of the peaks obtained from two different sized masks, one can assess the spread of the underlying edge. This is more reliable than trying to characterise the shape of the peaks, and allows an assessment of fuzziness which uses only the ability to find peaks and measure their amplitudes.

The amount of fuzziness associated with an edge may be characterised in two steps. Firstly, one finds the size of mask at which the edge ceases to appear like a step function: this can be recognised by comparing the amplitudes of the values obtained with successively smaller masks, and it corresponds roughly to that region of the spatial frequency spectrum within which the important information characterising the type of edge will be found. The second step is to code the relative amplitudes of the peak sizes due to masks of about that size. The order of magnitude of the result is more important than its exact value: the particular measurement that we use is da/b, where d is the panel width of the

smaller mask, and a and b are the peak amplitudes due to the larger and smaller masks. This provides a description that is satisfactory for many purposes. We suffer from the practical limitation of being able to use only two mask sizes at once, and this forces us to take special steps to recognise soft shading, whose principle energy may be concentrated in the longer frequencies (see e.g. the analysis of figure 7).

As well as its computational merits, the policy of deriving the low-level symbolic description from the smallest masks that give a measurable signal has psychophysical support. For example, in L.D.Harmon's well-known coarsely sampled and quantized photograph of Abraham Lincoln (reproduced for example in Julesz 1971 p311), perception of the face is impossible unless the high frequencies associated with the discretization are removed. One's choice of an operating region in the spatial frequency spectrum seems to be firmly involuntary. This limits the extent to which the computation of a rough, overall description of an image (which may be an important early stage of recognition), can rely on looking at the image through large masks.

Finally, to describe the representation of the result of the measurement, the modifier FUZZINESS is used, with the numerical value defined above. This number could of course be replaced by a qualitative descriptor, and will need to be converted to symbolic form before being passed to procedures that specialise in the shape of curved surfaces. Sharp edges have the associated modifier SHARP, and many object boundaries turn out to be sharp. Shadow boundaries have small fuzziness values, and those due to gradual curves of the underlying surface often

have quite large values. The examples given later illustrate these points.

The analysis of bar-mask convolutions

Sharp edges are easy to recognise using the criterion of sharp isolated peaks in the edge-mask convolutions, but the analysis of bar-mask profiles introduces more complex issues. Firstly, the possible-peaks are found in the two profiles (e.g. from bar-masks of panel widths 1 and 2), and they are matched. As before, except in special circumstances, peaks in neither record survive unless they find matches in the other. The exceptions are designed to deal with the case where peaks in the record from the smaller mask are well-defined and not small, but are closer together than can properly be resolved in the larger mask's record. This circumstance can occur, for example, when an edge of small amplitude occurs very near one of large amplitude and the same sign.

Before the pairs of peaks from the two sizes of mask may be parsed, it must be checked that they satisfy the isolation criterion. Accordingly, the pairs are arranged into disjoint groups such that each member of one group is at least 3d from a member of any other group. This constraint avoids the boundary condition problems described by Marr (1974a). If a group contains only one or two peak-pairs, it is ready to be parsed. If it contains more, one can sometimes split it by searching for typical edge configurations. These satisfy the following conditions: (i) They contain two peaks, of opposite sign, such that the amplitude of

one is greater than half that of the other. The reason for the half is that this circumstance cannot occur if the underlying image configuration is a thin bar. (This test provides another example of the use of conservative but reliable constraints to compute the description.) The actual numerical test applied uses the safe figure of 0.55, because one needs to allow for noise in the measurements.

(ii) The separation of the two peaks in the smaller bar mask's record does not exceed the separation of the peaks in the larger one's record.

These two criteria are frequently successful in breaking up large groups into their constituents (see figure 7 at x=724 for an example). If, after the application of both grouping procedures the remaining groups are still larger than 3, the resolution of the system has proved insufficient to characterise the image successfully at that point, and our implementation calls the result a GRATING. A few simple parameters are computed - like the width, the number of peaks (which equals the number of edges plus 1 in the ideal case), and the average intensity change associated with the peaks. It can happen that on successive nearby passes across the image, the low-level analysis hovers between a grating and a properly resolved description (see figure 8). (Algorithms that glue the very low-level assertions together can be made aware of this trouble, carrying descriptions across GRATING assertions if they match on both sides of them.) This GRATING assertion should be distinguished from the larger scale descriptor that one might invoke to describe a grating pattern spread across the visual field. The computation of such a

descriptor is not a low-level operation, in the sense of this article.

Parsing an isolated group of peaks

Provided that not more than three peak-pairs are contained in an isolated group, one of the following possibilities will be satisfied.

1: Three peak-pairs

If there are three pairs of peaks in the group, they are labelled CENTRE, LEFT-SIDEBAND and RIGHT-SIDEBAND. Provided (a) that both sidebands represent peaks whose values are opposite in sign from that of the CENTRE, and (b) that neither sideband has greater than half the amplitude of the CENTRE, the group may be diagnosed as a BAR whose amplitude is that of the CENTRE peak. If condition (a) is violated, the group is treated as a combination of types 2 and 3 below.

Assessing the width and the fuzziness of the bar requires very careful consideration (cf figure 3d). In practise, the peak separation is the best indicator of fuzziness, because it indicates the distance in the image over which the intensity changes at each side of the bar take place; and the relative amplitudes are the best indicator of bar width, because the smaller bar widths produce more interference between the effects of the edges at each side of the bar. Finally, the position of the BAR is given as the position of the CENTRE. Figure 3c shows a thin, sharp BAR (at around x=770), and figure 4 (x = 657) shows a fuzzy BAR.

2: Two peak-pairs

If there are two peak pairs in the group handed to the parser,

Figure 4. Profile 2 (see figure 1) is accompanied by its bar and edge mask convolutions, for various values of the panel width. The low-level symbolic representation of this image, obtained from edge-mask convolutions with panel size 16 (E16), and bar-mask convolutions with panel sizes 16 (B16) and 8 (B8), is the following:

```
EDGE (POSITION 256) (AMOUNT 258) (FUZZ SHARP)
BAR (POSITION 657) (AMOUNT -62) (FUZZ  FUZZY) (WIDTH 24)
EXTENDED-EDGE
        (POSITION 750) (AMOUNT 27) (FUZZ 17) (WIDTH 30) (DIRECTION +)
LINE (POSITION 595) (AMOUNT 21) (FUZZ SHARP) (WIDTH 16)
LINE (POSITION 694) (AMOUNT 20) (FUZZ SHARP)
```

Notice that peak separation is a better indicator of fuzziness for a bar than the relative amplitudes of the peaks from the two panel sizes (cf figure 3d). The description ignores the slow changes in intensity that are present in the image. These are picked up by the method from edge mask results with a panel size of 8 (description not shown). The LINEs that appear above would be subsumed in the description of the slow changes (see later in the article, and figures 5 and 7).

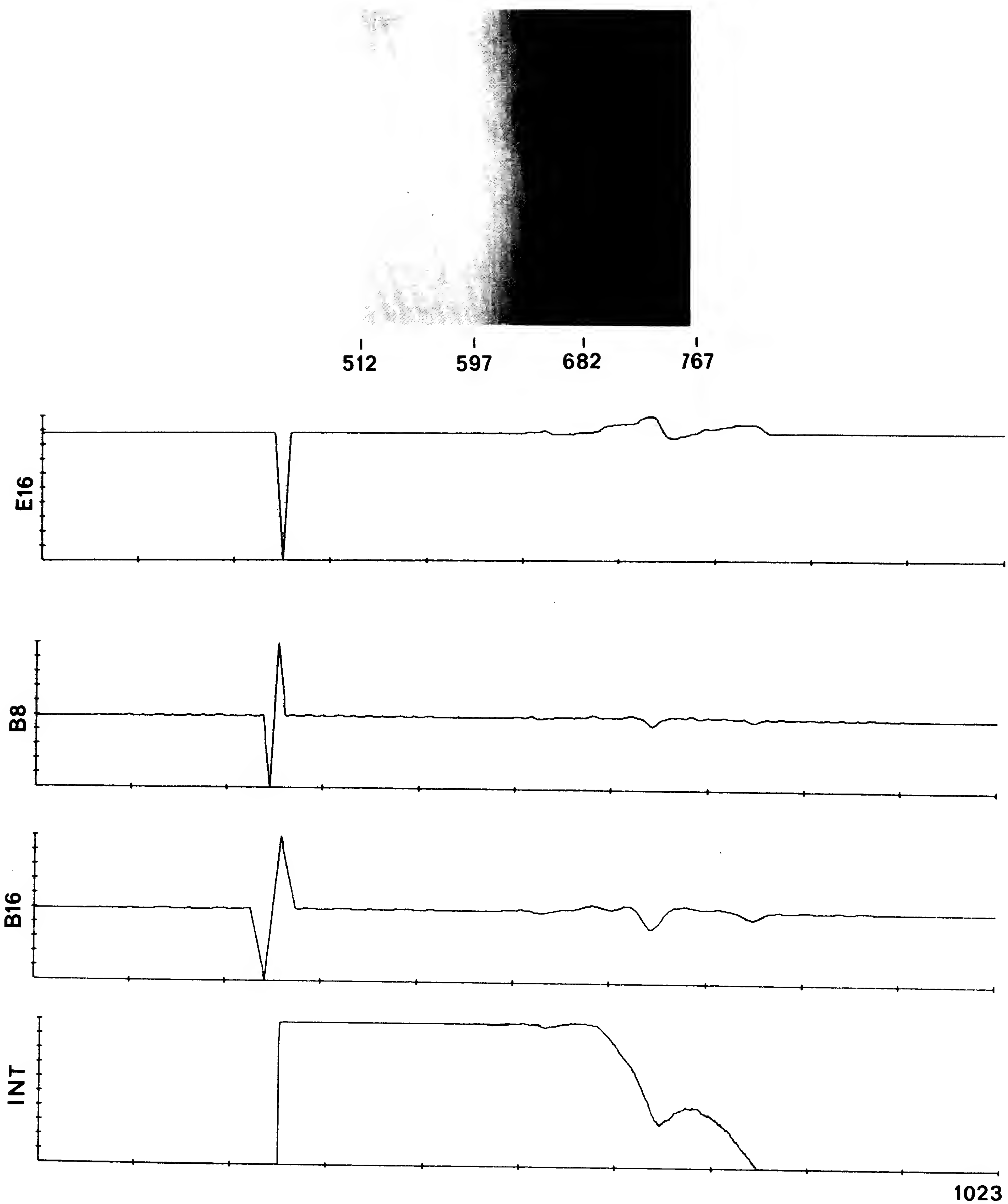


FIGURE 4

and the peaks are of opposite signs, the image contains an edge of some kind. If the peaks have the same sign, they are treated as two occurrences of case 3 below. The first category of edge is the classical one, where the amplitude of the smaller peak is greater than half (0.55) that of the larger. If the amplitudes of the peaks due to the two mask sizes are equal, and the separation of the peaks in each record is equal to the panel size for that record, the profile is that of a classical sharp edge. The edge's FUZZINESS is given as SHARP, its amplitude is the peak size, and its position lies mid-way between the two peaks.

If the amplitudes of the peaks due to the two mask sizes are not equal, the edge is described as being fuzzy by the appropriate amount. In such cases, the peak separation will also be greater than for a sharp edge. The amplitude of the edge is the amplitude currently being signalled at that point by an edge-shaped mask of appropriate size. The peaks in the record from the smaller bar-mask must not be significantly further apart than the peaks in the record from the larger one. If they are, the two mask sizes being used are probably too large, and details present in the image are being lost.

The criterion that the smaller of the two peaks in a record should be greater than 0.55 of the larger is an important one, because it signals that the EDGE description is appropriate. (It is used early on in the process to split up the groups of peak pairs). The question remains of what to do if there are only two peaks, but the amplitude of one is smaller than 0.55 of the amplitude of the other. There are two possibilities. The first one is that there is an edge of some kind in

the image, whose intensity change has started relatively gently, but finished abruptly. This situation is quite commonly produced by the shadow on a curved surface, and the associated edge is called an EXTENDED-EDGE. The second possibility is that there is in fact a BAR present, whose second SIDEBAND is missing or is too small to be seen, because part of the gradient change is very gradual on that side.

These two possibilities may be distinguished in the following way. If the image contains an EXTENDED-EDGE, the peaks in the large and in the small bar-mask records will occur at about the same place, because they correspond to genuine measurements of gradient change in the image. Accordingly, if this condition is satisfied, the parser assigns the term EXTENDED-EDGE to the configuration, and as well as the usual parameters, it is assigned a DIRECTION and a WIDTH. The FUZZINESS is computed in the usual way, by comparing the peak sizes in the two records. The WIDTH of the edge is obtained from the peak separation; and its amplitude, from the largest peak in the group. Figure 4 shows an EXTENDED-EDGE, and for comparison, figure 5 contains a fuzzy EDGE.

If, on the other hand, the peaks in the two records are roughly the distance apart of their respective panel-widths, the underlying image does not contain an EXTENDED-EDGE. In addition, one expects the ratios of the amplitudes of the CENTRE peaks of the two profiles to be larger than the ratios of the amplitudes of the SIDEBAND peaks. Using only the measurements of peak height, one cannot characterise the image more precisely, and so the parser calls it a BAR, with the usual parameters being assigned to it. This situation is extremely interesting, for in the

Figure 5. Profile 6 is a more complex distribution, containing several points of interest. Its analysis from an edge-mask of panel-width 8 (E8), and bar-masks of panel-widths 8 (B8) and 4 (B4), is as follows:

```
EDGE (POSITION 250) (AMOUNT 90) (FUZZ SHARP)
EDGE (POSITION 505) (AMOUNT -3) (FUZZ SHARP)
EDGE (POSITION 514) (AMOUNT 19) (FUZZ 4)
EXTENDED-EDGE
         (POSITION 615) (AMOUNT -9) (FUZZ 9) (WIDTH 10) (DIRECTION +)
EDGE (POSITION 634) (AMOUNT 2) (FUZZ SHARP)
EDGE (POSITION 645) (AMOUNT -15) (FUZZ 6)
EDGE (POSITION 675) (AMOUNT 1) (FUZZ 4)
EDGE (POSITION 733) (AMOUNT 14) (FUZZ 10)
EDGE (POSITION 766) (AMOUNT 1) (FUZZ 4)
EDGE (POSITION 815) (AMOUNT -66) (FUZZ SHARP)
EXTENDED-SHADING-EDGE (POSITION 691) (AMOUNT -4) (WIDTH 29) (START EDGE
675) (MIDDLE) (STOP LINE 718)
EXTENDED-SHADING-EDGE (POSITION 601) (AMOUNT -9) (WIDTH 58) (START LINE
550) (MIDDLE) (STOP EXTENDED-EDGE 615)
```

This interesting image contains both EXTENDED-EDGEs and fuzzy EDGEs. The edge at x=634, which was discovered by the program from the edge mask profile, gives rise to what is almost an illusion of extra lightness to its right (compare the intensity distribution). Notice that even very small intensity changes have been accurately described by the program.

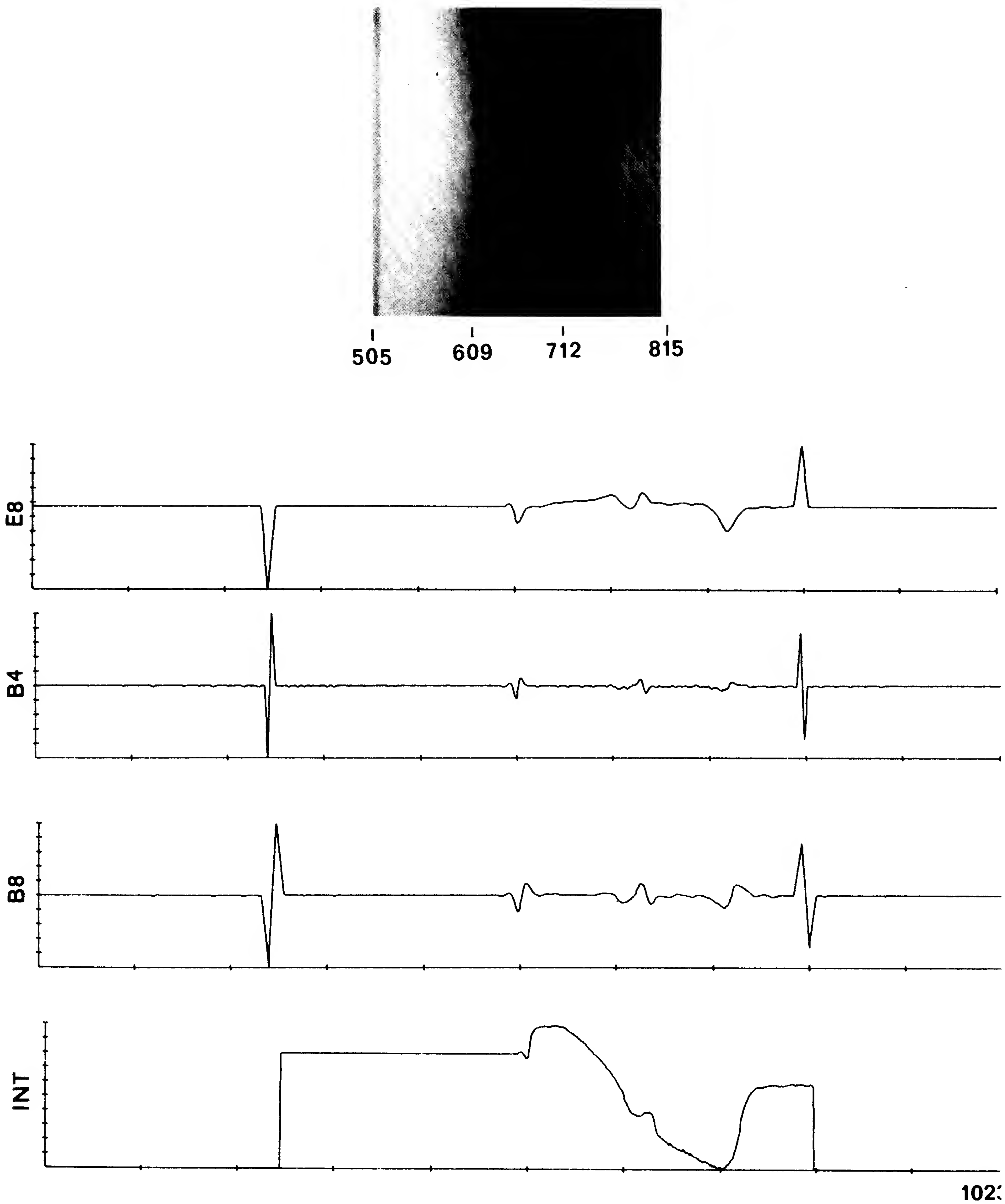


FIGURE 5

right circumstances it can give rise to a BAR assertion where we perceive a Mach Band. The criteria of peak separation and relative amplitude are what distinguishes an EXTENDED-EDGE from this type of BAR; and the difference between it and a normal EDGE is that the associated edge-mask peak is less sharp, and one of the two bar-mask peaks has less than half the amplitude of the other. Figure 6 shows a Mach Band whose BAR was obtained in this way.

One other point of interest is worth mentioning in connexion with figure 5: it is that the apparent lightness of the image (especially between coordinates 600 and 650) seems to be more closely related to the computed description of the intensity distribution than to the intensity distribution itself. This observation, and others (e.g. figure 6), suggest that one should look closely at methods for computing lightness that operate on a low-level symbolic description such as the one that is described here. Interestingly, methods for doing this are subject to simultaneous contrast phenomena if they are handed contrast measures of relative brightness, but treat them linearly as if they were straightforward measures of intensity change. In such circumstances, a method would for example tend to ascribe a greater apparent lightness to a grey square if its background were black than if it were white.

3: One peak-pair

Finally, there is the case where only one peak-pair is present in a group. This corresponds to places in the image where the second directional derivative is discontinuous, without any immediate change in

intensity. Such places are often important, because they can correspond to things like a crease on a(surface, or the nearer edge on a frontally) cube. Of course, one may be lucky and have some step change in intensity too, even if it is only a thin BAR caused by reflexions from the edge itself. The problem is how one should symbolize a change of intensity gradient: it is easy enough to recognise. It cannot be called an edge, because the strength associated with it would not reflect accurately the fact that there is no overall change in intensity at that point. Any intensity changes associated with it have to be symmetric, which commits one to coding it as a BAR of some kind. This is not unreasonable. Specular reflexions from sharp edges on an object can sometimes make them appear like very thin bars; and the accentuation of a boundary with a very thin line can make a painting look particularly realistic. Single bar-mask peaks are therefore coded as LINEs.

Although it is often sensible to treat LINEs as thin BARs, because they correspond to boundaries of objects or to segmentation points on a surface, such a description of the image will not accurately reflect the intensity distribution. The discrepancies correspond again to Mach Band illusions, but they differ from the failed EXTENDED-EDGE type because in this case, there is only one peak in the BAR mask profile. LINEs have to be quite strong before one perceives them as a band, but one certainly can.

Conventional thinking connects the Mach Band illusion to the centre-surround receptive field organisation of the retina, (see Ratliff 1965). There is no doubt that the measurement underlying the illusion is

Figure 6. The intensity profile contains an example of the Mach Band illusion. The parsing by the program using an edge mask of width 8 (E8), and bar masks of widths 8 (B8) and 4 (B4), is as follows:

```
BAR (POSITION 258) (AMOUNT -40) (FUZZ 6) (WIDTH 5)
EXTENDED-SHADING-EDGE (POSITION 269) (AMOUNT 20) (WIDTH 130)
        (START BAR 258)
```

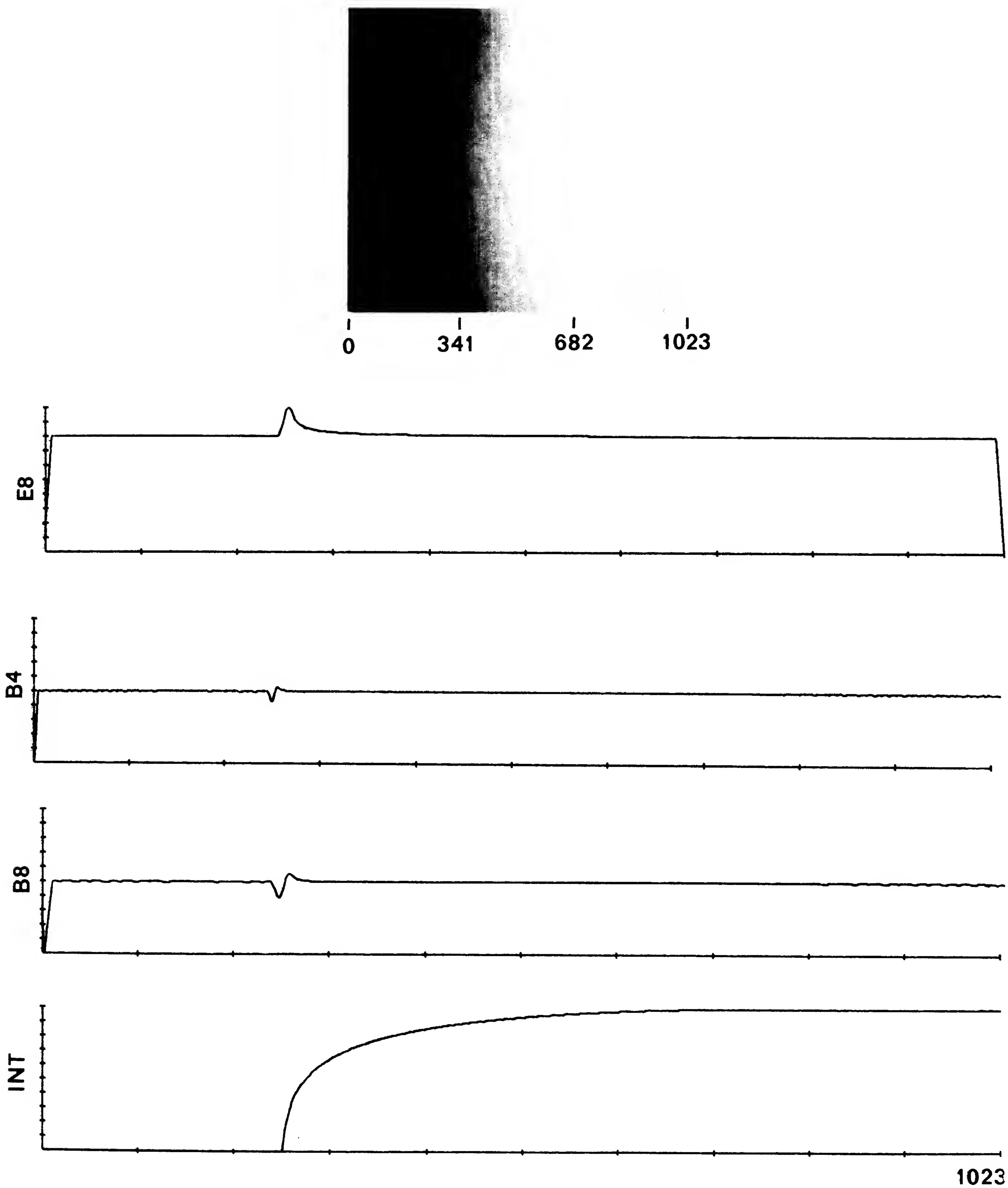


FIGURE 6

a measurement of the second derivative of intensity, but it is perhaps worth asking whether it might not be due more immediately to bar-shaped simple cells than to retinal ganglion cells. The present theory of low-level vision favours this view, and attributes the responsibility for the illusion to the mechanism that parses simple cell-like measurements of the second directional derivative of intensity into low-level symbolic descriptors. The existence of two distinct cases where the illusion may arise (a failed EXTENDED-EDGE, and a LINE), suggests that the particular implementation with which we are provided operates by trapping all of the alternative descriptions first, leaving BAR as a default for whatever bar-mask measurements remain.

The parser that was used to describe the images for this article keeps LINEs and BARs distinct. LINEs are fuzzy, because they correspond to the measurement of the second derivative (see earlier remarks about the relative amplitudes of bar-mask convolutions in these circumstances). They are assigned a width which is based on the smallest size of mask at which their presence is detectable (> 1% maximum value); but this width is not a well-founded measure.

If one uses only the methods and terms described so far, slow changes in intensity would tend to go unnoticed, even if they were quite large in amplitude. This may be seen in figure 4, whose representation fails to include a description of the slow intensity changes that

accompany the BAR. On some occasions, one may draw upon the measurements at a larger mask size in order to detect these changes, but in general, it appears that more immediately useful measurements arise from smaller edge-masks (measuring local gradient). (In figure 4, edge masks whose panel width is 8 provide the relevant measurements.) The reason why larger masks are of only limited help is that there is often a sharper intensity change nearby. Hence the large mask measurements fail to satisfy the isolation criterion, and their peaks will be a misleading indicator of the changes present in the image. For this reason, one has to use a high resolution analysis of gradient, and I summarise how slow changes are detected and described in our present implementation.

To detect slow changes in intensity, the program goes back to the edge-mask convolution. It splits the convolution into segments, by finding connected regions in which the result is either always positive or always negative. Small segments, and segments that correspond to items that have already been described, are removed. The remaining peaks are found, and those segments whose peaks have small amplitudes are ignored. The survivors correspond to items in the image that have not been dealt with properly elsewhere, and which cannot be ignored because their amplitudes indicate that something of note is present. The peak position, the peak value and the segment length are computed for each one, and wherever possible, LINEs and EDGEs are associated with a segment's beginning, middle, or end. These last associations are particularly important, because the start of gradual intensity changes is often picked up as a faint LINE by the central parser, and this allows

important features of the intensity change to be located precisely. In such cases, the LINEs in question are expunged from the main description.

The segment is classified as an EXTENDED-SHADING-EDGE, and the information listed above is included in its description. Such an edge may be thought of as an EXTENDED-EDGE, except that the DIRECTION parameter is unavailable. Figure 7 includes examples of such an edge.

Running the parser on a two-dimensional image

In order to illustrate the application of the method to a real image, figure 8 shows the results of running the whole process on a fairly complex image at two orientations. Isolated assertions (i.e. assertions that could not be glued to at least one neighbour at the appropriate orientation) would normally be ignored, but they have been included here to give a true idea of what the method finds in an image. LINEs that would normally be deleted because of their association with an EXTENDED-SHADING-EDGE, have also been included in the figure. The sensitivity of the process is not much in doubt: the very small circular indentations receive a fairly good analysis (the other orientations are missing here); and the very faint horizontal edges in the centre of the picture (at y = 73 and 75) have been noticed easily without, of course, the use of any high-level knowledge. The sensitivity can be set very high, because the parsing routines must be satisfied about a number of qualitative features of a profile before they make a choice about its description. interest, because it was recognised by the EXTENDED-SHADING-EDGE routines.

<u>Figure</u> <u>7</u>. Profile 5 (see figure 1) is similar to profile 6 (figure 5), but contains extra features. The description, obtained using an edge-mask of panel-width 8 (E8), and bar-masks of panel-widths 8 (B8) and 4 (B4), is as follows:

```
EDGE (POSITION 180) (AMOUNT 136) (FUZZ SHARP)
EDGE (POSITION 312) (AMOUNT 3) (FUZZ 4)
EDGE (POSITION 392) (AMOUNT 2) (FUZZ SHARP)
EDGE (POSITION 535) (AMOUNT -3) (FUZZ 4)
EDGE (POSITION 544) (AMOUNT 25) (FUZZ 5)
EDGE (POSITION 564) (AMOUNT 2) (FUZZ 4)
EDGE (POSITION 590) (AMOUNT 1) (FUZZ 4)
EXTENDED-EDGE
        (POSITION 682) (AMOUNT -12) (FUZZ 9) (WIDTH 14) (DIRECTION +)
EDGE (POSITION 724) (AMOUNT -20) (FUZZ 6)
EDGE (POSITION 776) (AMOUNT 3) (FUZZ 4)
EDGE (POSITION 784) (AMOUNT -4) (FUZZ 4)
EXTENDED-SHADING-EDGE (POSITION 670) (AMOUNT -14) (WIDTH 67)
        (STOP EXTENDED-EDGE 682)
EXTENDED-SHADING-EDGE (POSITION 491) (AMOUNT 4) (WIDTH 36)
        (START LINE 486)
EXTENDED-SHADING-EDGE (POSITION 439) (AMOUNT -8) (WIDTH 73)
        (START EDGE 392) (MIDDLE LINE 444)
```

Notice once again the detail that the method has described.

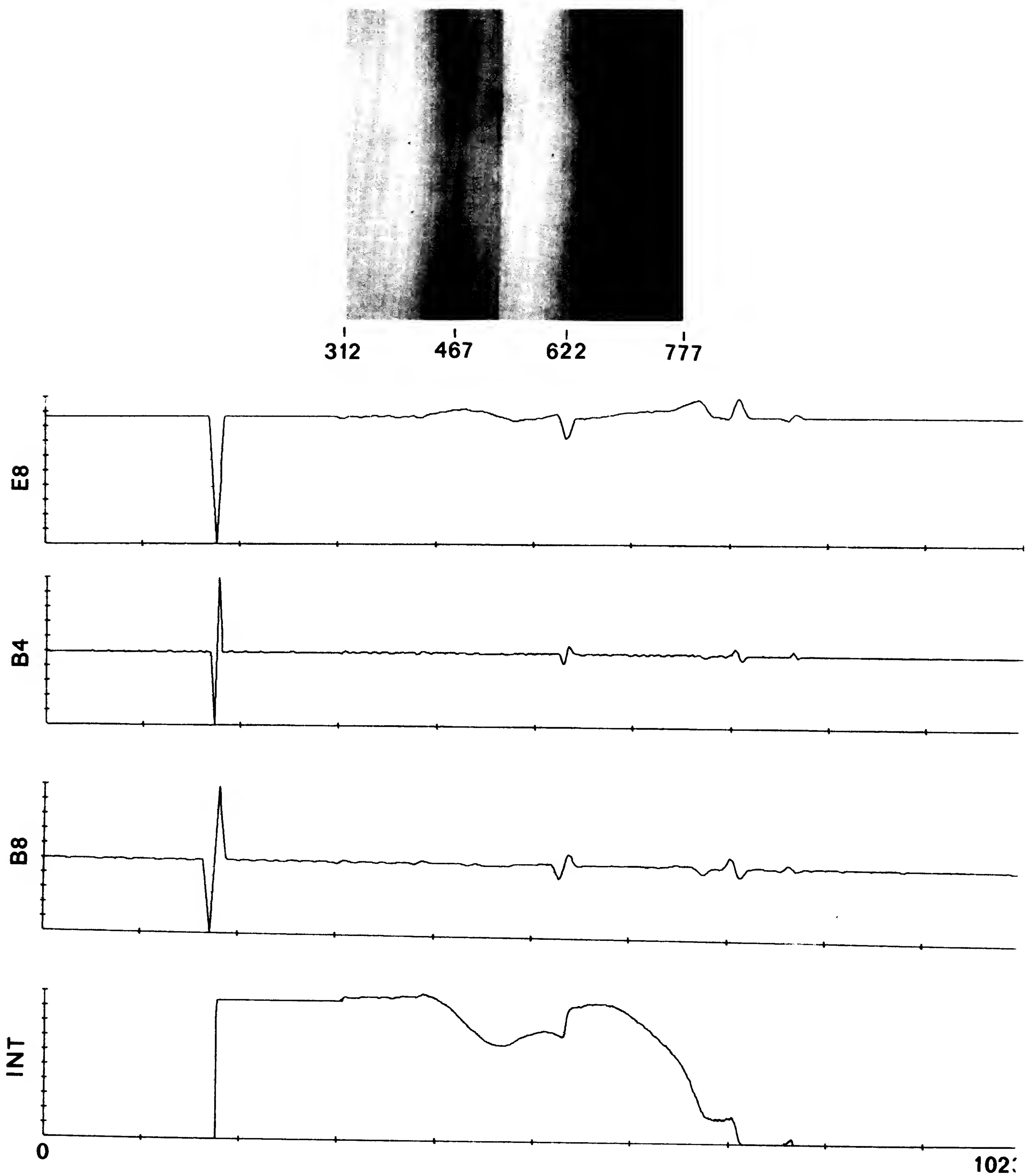


FIGURE 7

The dark shadow in the centre of the image is of interest, because it was recognised by the EXTENDED-SHADING-EDGE routines. It extends over a considerably larger region than the panel widths of the masks (which were 1 and 2 for the parsing that is shown). There are certain aspects of the EXTENDED-SHADING-EDGE process that are unsatisfactory: this is partly because we are forced to use only two sizes of mask, and partly because very extended edges are too spread out to be dealt with entirely at this low level. They need to be treated almost as if they were a local texture (Marr 1975).

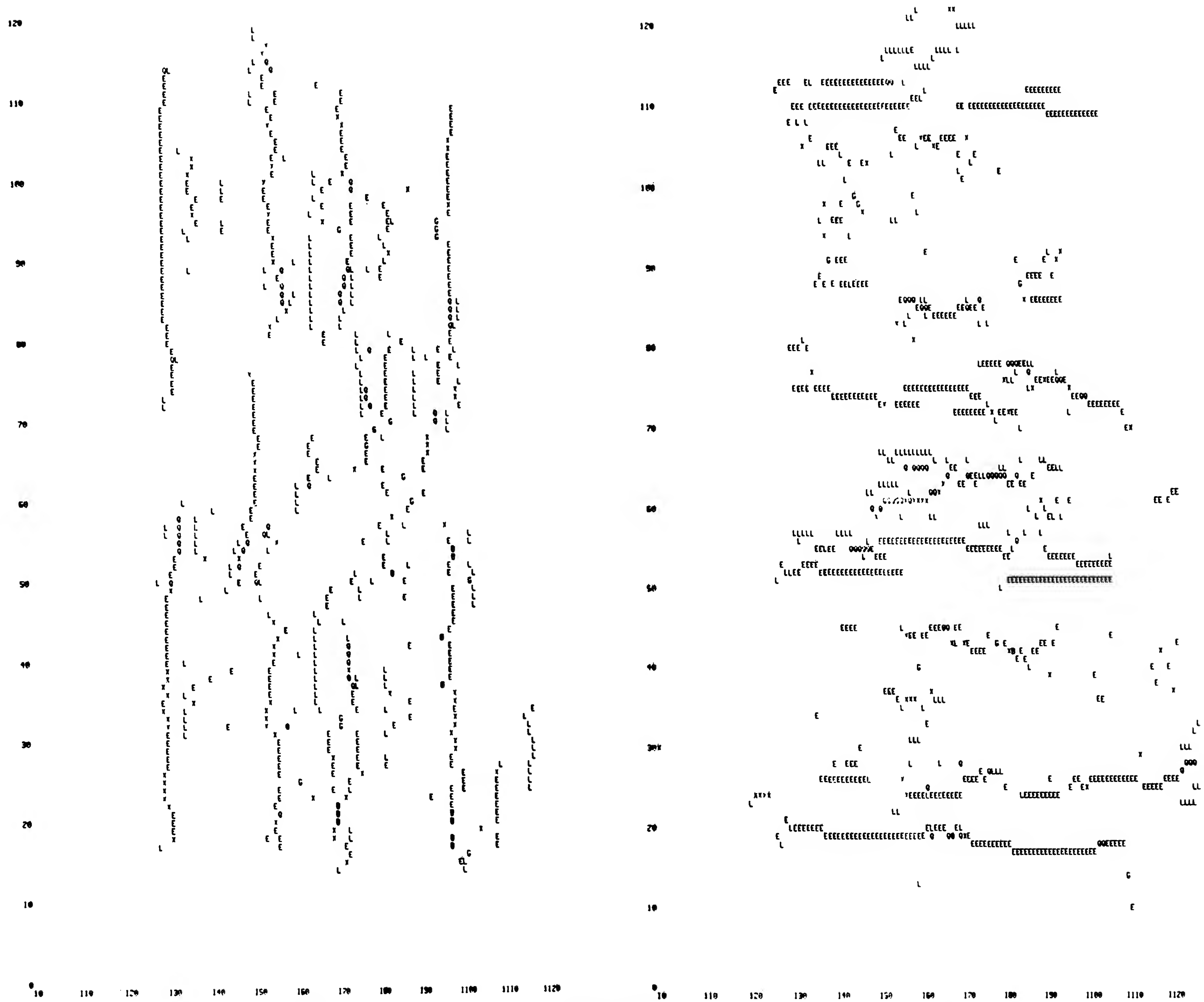
Finally, the reader will have noticed that a number of issues arise when one contemplates the interaction of information at different orientations. For example, should the interaction take place before or after parsing? What are the rules for carrying it out, and how are they arrived at? These are important questions, whose answer is not straightforward, and they will be dealt with elsewhere.

Completeness

The theory behind this article is that the purpose of low-level vision is to compute a very low-level symbolic description of the intensity changes in an image, which is sufficiently expressive that subsequent processes need have access only to this description (Marr 1974a). There is a sense in which this process resembles an inverse of the original measurements, and one can therefore ask how faithfully the image is described. We have already seen that the process is not

Figure 8. To illustrate the application of the process to a two-dimensional image, horizontal (8a) and vertical (8b) parsings have been computed for the image that appeared in figure 2. The assertions have been represented in by the following conventions: E = edge, L = line, X = extended-edge, Q = extended-shading-edge, and G = grating. The full parsing along for example the line x = 58 is the following:

```
EDGE (ORIENTATION 4) (POSITION 58 19) (AMOUNT 43) (FUZZ 1)
EDGE (ORIENTATION 4) (POSITION 58 24) (AMOUNT 26) (FUZZ 1)
EXTENDED-EDGE (ORIENTATION 4) (POSITION 58 36)
        (AMOUNT -10) (FUZZ 1) (WIDTH 2) (DIRECTION -)
EDGE (ORIENTATION 4) (POSITION 58 44) (AMOUNT -7) (FUZZ 1)
EDGE (ORIENTATION 4) (POSITION 58 56) (AMOUNT -112) (FUZZ SHARP)
EXTENDED-EDGE (ORIENTATION 4) (POSITION 58 61)
        (AMOUNT -34) (FUZZ 2) (WIDTH 3) (DIRECTION -)
EXTENDED-SHADING-EDGE (ORIENTATION 4) (POSITION 58 65)
        (AMOUNT 20) (WIDTH 4) (STOP LINE 67)
EDGE (ORIENTATION 4) (POSITION 58 73) (AMOUNT -7) (FUZZ 1)
EDGE (ORIENTATION 4) (POSITION 58 75) (AMOUNT 8) (FUZZ SHARP)
EXTENDED-EDGE (ORIENTATION 4) (POSITION 58 81)
        (AMOUNT -19) (FUZZ 2) (WIDTH 6)
EXTENDED-SHADING-EDGE (ORIENTATION 4) (POSITION 58 86)
        (AMOUNT -22) (WIDTH 9)
EDGE (ORIENTATION 4) (POSITION 58 99) (AMOUNT -2) (FUZZ 1)
EDGE (ORIENTATION 4) (POSITION 58 111) (AMOUNT -9) (FUZZ 1)
EDGE (ORIENTATION 4) (POSITION 58 117) (AMOUNT -6) (FUZZ 1)
{LINE (ORIENTATION 4) (POSITION 58 67) (AMOUNT 16) (WIDTH 1)
 LINE (ORIENTATION 4) (POSITION 58 121) (AMOUNT 7) (WIDTH 2)
 LINE (ORIENTATION 4) (POSITION 58 31) (AMOUNT -6) (WIDTH 2)}
```



(a) FIGURE 8 (b)

sensitive to changing the image at isolated points, but this is not a disadvantage if one assumes, as we are, that the interesting intensity changes occur over groups of several image points. (It also protects the system to some degree against white noise). Nor is the process sensitive to changes in the image that cause changes in the shapes, but not the positions or amplitudes, of peaks in the mask-response profiles. It is however very difficult to produce a change in the shape of a peak in one mask's convolution profile that does not affect the size of the peak in the profile obtained from a mask of a different size.

The question was raised in the introduction of whether the family of descriptors introduced here provides a fine enough covering to allow adequate shape discrimination based on shading alone: until the later programs are completed, there is no satisfactory way to test this. Of the inverse property one can however be more confident: provided that there is a sufficiently isolated change of intensity, or of intensity gradient, that involves several nearby image elements, it will show up in the convolution profiles, and will therefore be described in some way by the subsequent parsing process. If the intensity change is sharp and not too small, one can relax the isolation condition (Marr 1974b).

This article sought to establish the following points: firstly, that edge- and bar-mask convolutions with an image may profitably be thought of as measuring the first and second directional derivatives of

intensity in an image. Secondly, that the process of interpreting such convolutions is not trivial. Thirdly, that they may be interpreted by converting the measurements immediately into a low-level symbolic description of the intensity array. Fourthly, that although specular reflexions and certain other kinds of intensity change are not treated here, it is already apparent that this kind of description does not require a great number of primitives. Fifthly, that it can be accomplished by methods which use only rather simple features of the original measurements, like the sizes and positions of peaks, and whether they are positive or negative. And sixthly, that a parsing algorithm based on the methods described here runs about as well as could be expected on natural images. In addition to these points, it is noted that a full explanation of the Mach Band illusion must include an account of the relationship between measurements of the second derivative of intensity, and the symbolic interpretation of those measurements.

Within the framework of the method itself, certain operations may be isolated as being of especial importance. The following perhaps deserve special mention: the use of different mask sizes for the detection of peaks in the convolution profiles; the comparison of peak sizes and positions obtained using those different mask sizes; the precedence of certain peak configurations (for example the classical EDGE peak-pairs) and their usefulness in decomposing larger groups of peaks; and the importance of using only conservative and well-founded procedures at all stages during the analysis. This last point requires a sensitivity to hidden issues, like those that concern the boundary conditions of the

related inverse transform. Finally, it is important for a vision system to have an adequate range of mask sizes available: this feature is unfortunately extremely costly to implement in a general-purpose computing installation, though it seems to be available in advanced mammalian visual systems.

The broader computational justification for this approach to vision will rest upon the extent to which a vision system that uses the low-level package defined here actually works. This question is taken up elsewhere (Marr 1975).

Acknowledgements: I thank Ken Forbus, whose enthusiasm and skill in helping me implement the methods described here were very valuable. The study could not have been carried out without the extensive and flexible computer facilities that are available at the Artificial Intelligence Laboratory. Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract number N00014-70-a-0362-0005.

References

Cooley, J.W. & Tukey, J.W. (1965). An algorithm for the machine computation of complex Fourier series. Math. Comp., 19, 297-301.

Cornsweet, T.N. (1970). Visual perception. New York: Academic Press.

Hubel, D.H. & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (Lond.), 160, 106-154.

Maffei, L. & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. Vision Res., 13, 1255-1267.

Marr, D. (1974a). The purpose of low-level vision. MIT Artificial Inteligence Laboratory Memo 324.

Marr, D. (1974b). The recognition of sharp, closely spaced edges. MIT Artificial Intelligence Laboratory Memo 326.

Marr, D. (1975). Configurations, regions, and simple texture vision. (In preparation).

Ratliff, F. (1965). _Mach Bands: quantitative studies on neural networks in the retina_. San Francisco: Holden-Day.